A Real Student Network Analysis and Mining in Class Teaching

Xian Hua Zeng*, Sheng Wei Qu, Zhi Long Wu, Xu Cheng and Jing Jing Nie

Department of Intelligence Science and Technology, College of Computer Science and Technology, Chongqing University of Posts and Telecommunications,Chongqing400065,China

Email:xianhuazeng@gmail.com

Keywords: Education; Student-student interrelation networks; Directed Graph; Visualization.

Abstract. The educational data has a wealth of information. In this paper, we construct Student-Student Interrelation Networks (SSIN) to analyze and find meaningful information from student relation data during class teaching. We collect a group of data about students in a new teaching class, then we use adjacency matrixes to represent student-student interrelations. In order to have a better visualization, we construct the directed graph with respect to adjacency matrixes. Then we analyze community structures and their hiding information. It is benefit for improving teaching & learning activities and make related plans for guiding student development.

Introduction

Data mining and machine learning techniques are the field of discovering novel and potentially useful information and knowledge from large amounts of data [1]. Recent advances in data mining and machine learning methods have shown a series of successful applications across a wide variety of fields such as science and education. Education is one of the most powerful instruments for human development to a country. Educational data is exceedingly rich and may come from educational questionnaire survey, learning management systems, interactive learning environments, intelligent tutoring systems, etc. Educational data mining and analysis can help teachers enhance teaching and improve students' learning efficiency and methods[2-6]. In 2012, U.S. Department of Education Office of Educational Technology delivered the government report of 'Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics'[7]. Educational data mining can help educators communicate with students and parents about the found information and help teachers be more effective in the classroom with more real-time and making recommendation services. Of course, some higher education institutions are using analytics for improving the services for increasing student scores and abilities.

Student relational data is a kind of general and convenient educational data. This paper analyzes the student-student interrelationship which is from at Grade 2012 Undergraduate Teaching Class, Department of Intelligence Science and Technology, College of Computer Science and Technology, Chongqing University of Posts and Telecommunications in China. Two student-student interrelation network graphs are constructed at interval of ten weeks. Some community structures and changing structures including this two networks are analyzed and studied. According to the community structure, students' learning styles, history, academic performance, school conditions and hobbies are analyzed and mined. We try to apply these information helping students such as correcting their learning behaviors, improving study efficiency, making the training plan and training professional ability.

Educational Data Collection and Description

The data for our analysis was from questionnaires in our class teaching of Artificial Intelligence course. Our survey taken at Grade 2012 Undergraduate Teaching Class, department of Intelligence Science and Technology, College of Computer Science and Technology, Chongqing University of Posts and Telecommunications in China. The data was gathered and prepared by Dr. Xianhua Zeng.

Two questionnaires was respectively carried in the first week and the eleventh week. The sixty-six attendees were asked to answer their most familiar five-classmate during the teaching of Intelligence course. All attendees who selected Intelligence Science and Technology major during the third and fourth study-year, are grouped into a new teaching class after enrolment and training in large categories. In the first week, each attendee were asked to fill their most familiar five classmates. In the eleventh week, each attendee were asked to give each of their most familiar five classmates a score at interval 1-5, where 5 denotes the most familiar score. During two questionnaires, four students did not return questionnaire data so we completely removed their information, and sixty-two participants responded and gave us permission to make their data available in our anonymized dataset.

Data Structure and Mining Method

Adjacency Matrix: We denote each participated student a node, so sixty-two participants correspond sixty-two nodes, where node Si has a value to node Sj if person Si reports that person Sj is one of his or her the most familiar five, i.e. formatting a matrix S, called as the adjacency matrix. Each row is from the data of each participant. To the questionnaire data in the first week, a sparse square matrix that acts as a connection matrix, that is, a value of 1 indicates a connection between nodes while 0 indicates no connection. The first n rows/columns is equal to the number of valid participants, and the n+1th row denotes the in-degree of each node, as shown Table1. In the 11-th week, the other sparse square matrix that acts as a connection matrix, that is, a value between 0-5 indicates a familiar degree between participants, and the last row denotes the in-degree of each node, as shown Table2.

Student	S1	S2	S3	S4	S5	…
S1	0	0	1	0	0	…
S2	0	0	0	0	0	…
S3	1	0	0	0	0	…
S4	0	0	0	0	0	…
…	…	…	…	…	…	…
S62						
In-degree	6	4	6	4	5	

Table1 the questionnaire data in the first week

Student	S1	S2	S3	S4	S5	…
S1	0	0	4	0	0	…
S2	0	0	0	0	0	…
S3	1	0	0	0	0	…
S4	0	0	0	0	0	…
…	…	…	…	…	…	…
S62						
In-degree	17	15	22	21	18	

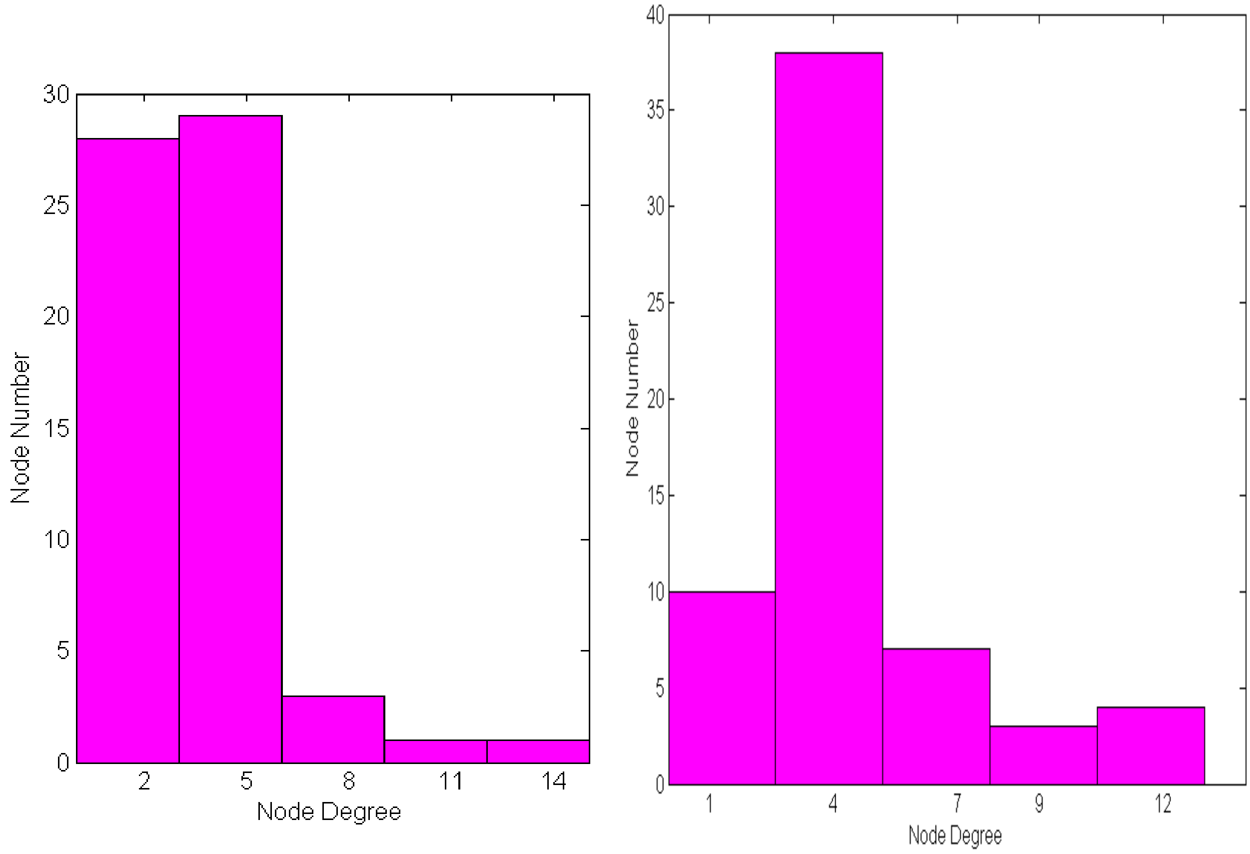
Table 2 the questionnaire data in the eleventh week

Directed Graph(Student-Student Interrelation Network, SSIN). We construct two connected networks to represent two student-student interrelation networks from two adjacency matrixes with respectively corresponding to the questionnaires of the first week and the eleventh week. The first network, i.e. the directed graph, includes 245 edges between 62 nodes with responding to the first questionnaire, and the second network includes 283 edges between 62 nodes with responding to the second questionnaire. After ten-week class teaching activity, 38 edges are added into the second student-student interrelation networks which show to build the relations between 32 student-pairs.

Visualization Method. This paper applies the direct graph theory algorithms[8] to visualize student-student Interrelations. A biograph object is a data structure containing student-student interconnected data used to implement a directed graph. Nodes represent students, and edges represent interrelationships between the nodes. This biograph object also stores information, such as color properties and text label characteristics, used to create a 2-D visualization of the graph. In our experiments, we use a simple graph visualization tool, i.e. biograph.m function in Matllab2013. We assume the graph is represented as an adjacency matrix used to create a 2-D visualization of the graph. The 2-D visualizations of student-student interrelations can view a graphical representation of a biograph[8] object using the view method.

Visualization and Analysis of SSIN

Figure1(a) shows the degree distribution of the first network with responding to the first questionnaire (including 62 participants) in the first week of class teaching, which also exhibits heterogeneous degree distribution features as the maximum degree is 16 while the smallest is only 0. Figure1(b) shows the degree distribution of the second network with responding to data after tent weeks of class teaching, which also exhibits significant heterogeneous degree distribution features as the maximum degree is 14 while the smallest is only 0.



(a) Histograms of Node Degree in the first week (b) Histograms of Node Degree in the 11th week

Figure.1 the degree distribution of two student-student interrelation network which are obtained by two questionnaires of the first week and the 11th week.

We use the graph visualization tool (biograph.m in matlab) to respectively exhibit the visualization figures of two student-student interrelation networks with corresponding to two group questionnaire data in the first week and eleventh week of teaching class. The detail analysis features are marked in Figure. 2 and Figure. 3.

Figure. 2 show the six participant-groups and one outlier in the first week of class teaching. Small community structures are very significant. Group1, 2 5,6,7 are respectively made up of students from the same original classes. Group 3 is made up of students from near living places, especially in the same dormitory, and Student26 has good leadership and communication ability. Of course, Student43 is seemly a bit of loneliness and need to be cared.

After eleven weeks, the interrelation between sixty-two students has significantly changes, as shown in Figure 3. The new visualization figure shows that this teaching class has developed into two bigger categories. Students from the same category have much similarity. There are many common characteristics in x1, x3. The students in x2 has developed the same activity group, while students in x4 and x5 are in the other activity group respectively. Especially, student-student interrelations have become more and more close classmate relation while some significant and stable community structures are emerging

Two visualization figures clearly exhibit the relationship and other habits between these students, and the changing after eleven weeks. Teachers can understand the students' characteristics with these information and come up with a better teaching plan. At the same time, there is also a better understanding among students.

Conclusion

There is much helpful information in educational data, among which the student-student interrelation is wealthier. In this paper, we mine and analyze the data from the student-student interrelation. We use matrixes to record these relationships, and directed graphs to visualize them. Through the experiments, we find much implicit information. With the analysis, these mined information can help students improve their efficiency and study behaviors, or help them make study schedules. At the same time, we have a better understanding and application about education data mining.

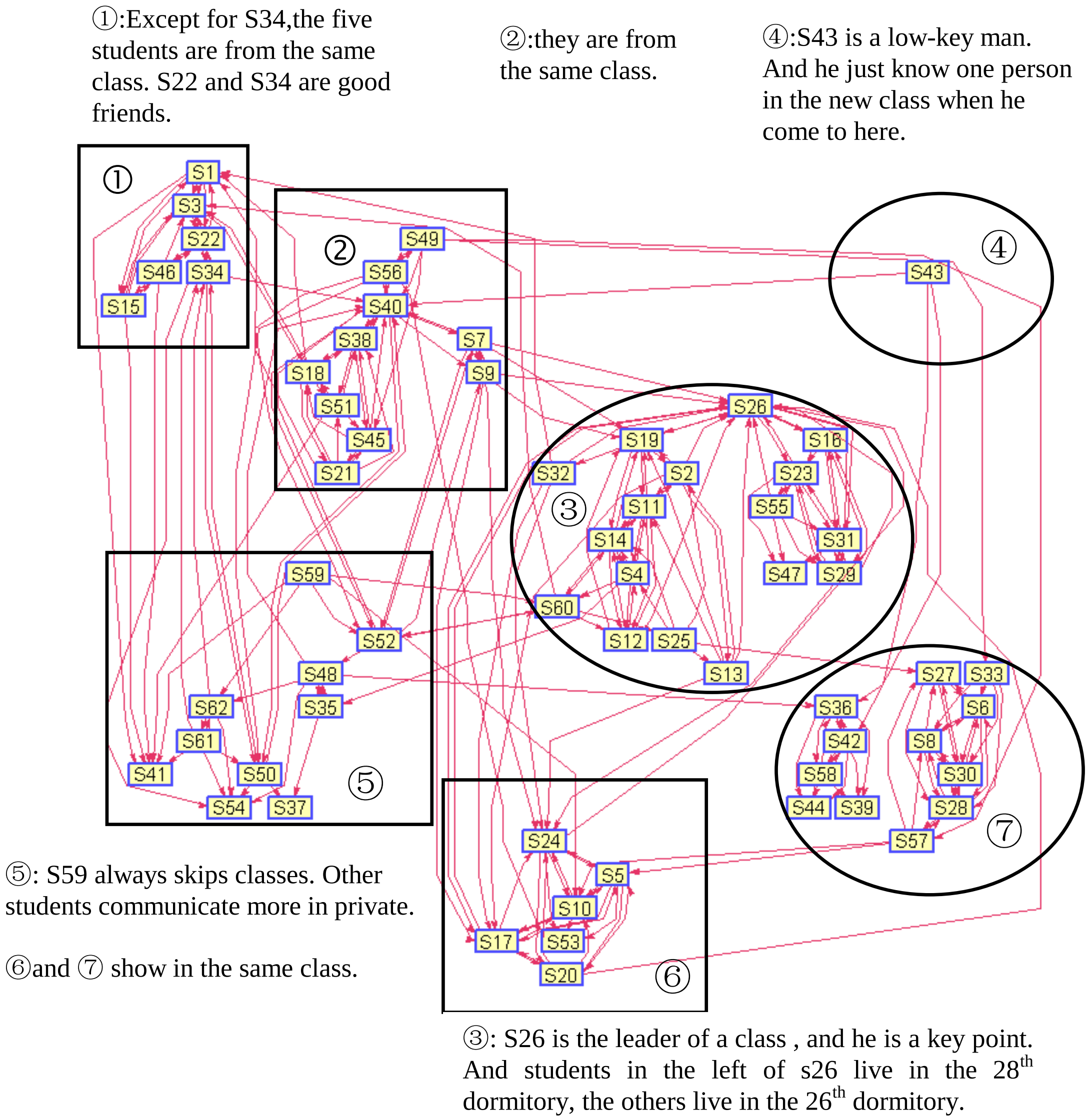


Figure. 2 Visualization and analysis of the student-student interrelation network in the first week of a new teaching class.

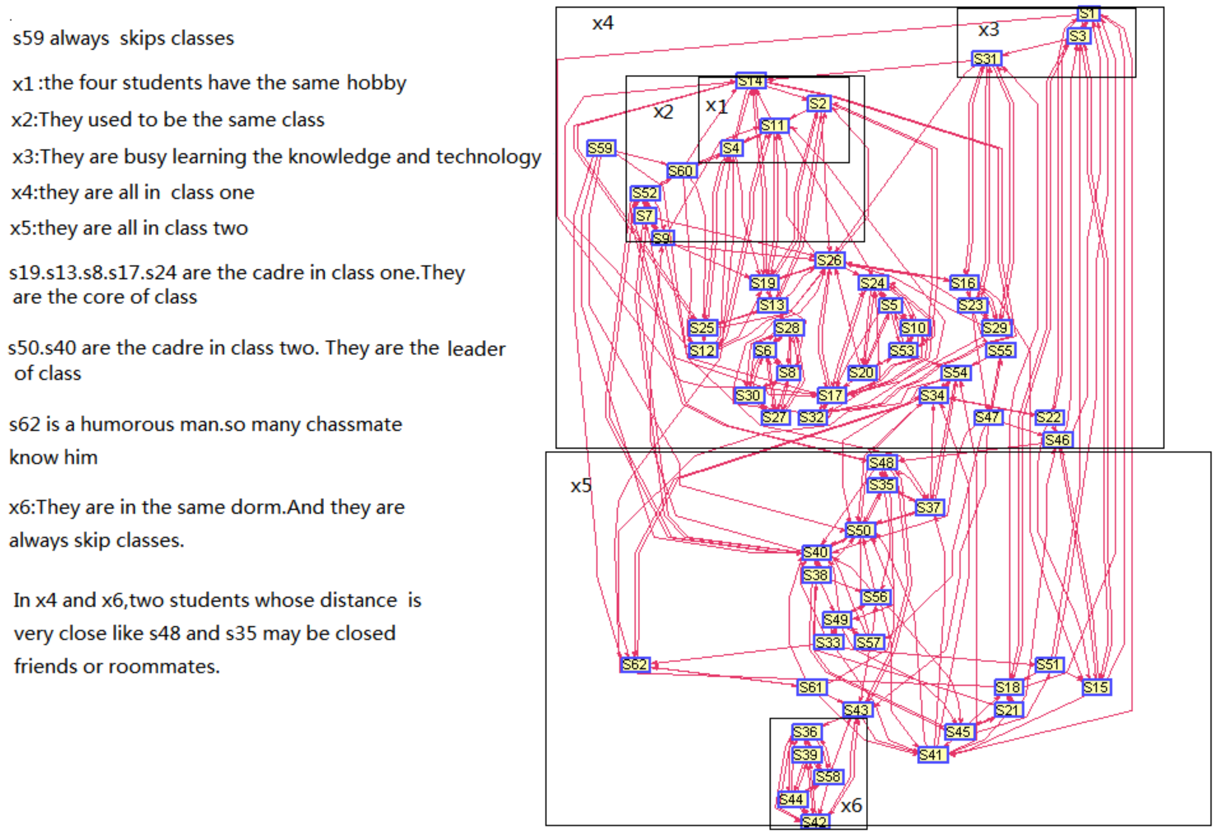


Figure. 3 Visualization and analysis of the student-student interrelation network after ten weeks of the new teaching class.

References

[1] E. Mjolsness, D. DeCoste, Machine learning for science: state of the art and future prospects[J].Science,2001(5537): 2051-2055.

[2] M. Bienkowski, M. Feng, B. Means, Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics, U.S. Department of Education Office of Educational Technology,2012.

[3] Y. Bergner, Z. Shu and A. Davier. Visualization and Confirmatory Clustering of Sequence Data from a Simulation-Based Assessment Task . EDM 2014: The Seventh International Conference on Educational Data Mining, pp 177-184.

[4] E. Watanabe, T. Ozeki and T. Kohama. Analysis and extraction of behaviors by students in lectures. EDM 2014: The Seventh International Conference on Educational Data Mining, pp. 329-330.

[5] G. Rebolledo-Mendez, B. D. Boulay, R. Luckin, E. I. Benitez-Guerrero. Mining Data From Interactions With a Motivational-aware Tutoring System Using Data Visualization. Journal of Educational Data Mining, 2013,5(1)72-103.

[6] M. Desmarais, F. Lemieux Clustering. Visualizing Study State Sequences. EDM 2013: Proceedings of 6th International Conference on Educational Data Mining.pp.153-158.

[7] M. Eagle, M. Johnson, T. Barnes. Interaction Networks: Generating High Level Hints Based on Network Community Clustering. EDM 2012: Proceedings of the 5th International Conference on Educational Data Mining, pp.164–167.

[8] http://cn.mathworks.com/help/bioinfo/ref/dolayoutbiograph.html.